

Weekly Report (2018.10.8-2018.10.14)

DONE

1. 准备并报告组会报告

EmbeddingVis: A Visual Analytics Approach to Comparative Network Embedding Inspection 本文设计了一个探索性可视化分析系统，该系统支持在集群，实例和结构层面上可视比较与理解嵌入向量。更具体地说，它有助于比较不同嵌入模型中节点度量的保留内容和方式，以及调查节点度量与所选嵌入向量之间的关系。

2. RSATree

- 重构旧版代码，将新的内容融入到旧版代码中，把代码跑起来。小组分工，我和舒悦负责修改前端与后期测试。
- 由于 size 感知项目在准备正式实验前的讨论，确认需要继续优化实验，RSATree 部分工作暂时搁置。

3. size 感知投稿项目：

- 在准备正式实验的过程中发现：实验2中的变量半径的取值，还没有通过预实验进行验证；均匀分布与高斯分布的 bias 差异还没有进行对比，需要补做了均匀分布的预实验。通过预实验与结果分析，初步确定了以上两点，半径的取值范围以及均匀分布与高斯分布的 scale-bias 线性关系图的斜率不同。
- 准备正式实验：a. 修改前端代码，使整个实验过程完全自动化，不同参数的切换以及结果的保存，全程不需要人的干预。b. 整理前面所有预实验的结果。c. 整理实验流程，包括实验前的实验介绍、实验与案例说明、正式实验的试操作系统、疲劳程度的量化测试、实验中的实验顺序以及实验后的问卷设计。
- 正式实验前的讨论：a. 一同再次分析前面所有预实验的结果，发现实验结果非常不稳定，需要尝试阶梯法 trail 分配方法，来取代现在的随机分配 trail 题目的方式。调研连续做两次拟合对误差的影响。b. 对于整个实验流程，疲劳程度尽量做到量化测试，因为在预实验中发现疲劳程度对结果的影响很大，做了一些调研，还没看到量化测试疲劳程度的类似工作，然后考虑针对我们的实验来做这个疲劳程度测试，测试一组强度略大一点的实验，然后用正确率来估计一下，如果没有达标，要不就是没有理解我们的实验，要不就是太疲劳；实验中的实验顺序考虑使用拉丁方平衡顺序进行排序，使得每位被试的实验顺序不同，避免实验顺序对实验结果的影响；实验后的问卷需要进行完善，需要是（固定题目 + 开放性题目）这种模式。
- 实验问卷的设计：设计问卷，首先要思考我要分析什么，前面做的数据分析都是客观分析，而问卷中的选择题与聊天都是主观分析。问卷设计有两个思路：a. 做主观与客观的对比，需要思考客观分析实验结果，列出准备写进文章的几个点，比如：密度是否有 bias，bias 什么分布，那么主观问题就问：你觉得 xxxx 移动（变化）对密度感知是否存在 bias，如果存在 1-5 你觉得大还是小，如果存在 你觉得是递增还是递减。这样就利用主观问卷 让用户从头到尾把客观问卷准备用数据分析解释的事情来一遍，到时候论文就可以去寻找：主观 和 客观 一致或者不一致，都可以写很多分析。在这种思路下，就要思考 客观分析实验结果 到底需要 list out 多少 conclusions，来对应设计主观。b. 补充 客观分析搞不定的内容，思考哪些内容是 客观分析分析不到的人可能会有感觉的，依此来设计主观问卷。
- evaluation 实验的迭代真的是很多很多，在觉得应该 ok 的时候，又有很多需要改进的地方。

4. 与王杰学弟面谈：

后面我们会一起合作做项目，这位同学态度很积极，学习兴趣很高，算法、数据结构基础相对较强，做项目与代码经历不多，需要加强代码训练。目前是先让这位学弟去补足代码基础，了解前端，熟悉后端以及前后端通信，后面准备让他主要负责后端。

今天讨论到均匀分布与高斯分布的 bias 差异，发现 density 预实验全部是用高斯分布做的，补做了均匀分布的预实验，结果 scale-bias 不是理想的线性关系，明早重新做一遍。

2. 实验2中的变量半径的取值：还没有通过预实验进行验证，明早进行预实验。

3. 修补实验系统，实现实验流程自动化。

小结

工作日工作时长 $13.5+9.5+6+10.5+9.5$ ，周末5小时，总时长约54h。周三去杭口补了3颗牙，用了一下午的时间。

PLAN

短期计划（一周）

1. size 感知投稿项目：将实验整体流程换成阶梯法，量化测试疲劳程度，减少实验结果拟合误差，迭代完善实验问卷。
2. RSATree 投稿项目：把旧系统中的几个页面全部跑起来并做测试。
3. 调研动态图时间片截取相关工作，机器学习如何融入到工作中。

中期计划

1. Visevo论文：阅读 Data Illustrator, <https://www.cc.gatech.edu/gvu/ii/dataillustrator/>。
2. 动态图时间片截取项目：10月份需要进行充分的调研，学习涉及到的知识，包括TCP Tree数据结构，动态图相关的一些知识，机器学习如何融合等，搞清楚具体做什么，如何做，10月底开始写代码尝试进行尝试。

长期计划

1. 学习更多机器学习、数据挖掘相关的算法。
2. 在项目中锻炼自己的思考能力与代码能力。